

# A Central Limit Theorem on $GL_n(F_q)$

William M. Y. Goh,\* and Eric Schmutz\*†

Drexel University, Mathematics Department, Philadelphia, PA 19104

## ABSTRACT

For  $T \in GL_n(F_q)$ , let  $\Omega_n(T)$  be the number of irreducible factors that the characteristic polynomial of  $T$  has. We prove that, for any fixed  $x$ ,  $\#\{T: \Omega_n(T) < \log n + x\sqrt{\log n}\} / \#GL_n(F_q) \rightarrow (1/\sqrt{2\pi}) \int_{-x}^x e^{-t^2/2}$  as  $n \rightarrow \infty$ .

*Keywords:* limit theorem, asymptotics, cycle index, characteristic polynomial, finite field

Many asymptotic (large  $n$ ) results on  $S_n$  should have natural analogues on  $GL_n(F_q)$ . This viewpoint was advocated by Stong [12], who used Kung's vector space cycle index to prove several such theorems. He and Diaconis proposed a possible analogue of Goncharov's well-known theorem on the distribution of the number of cycles on  $S_n$ . We shall prove it; the main result in this paper is a central limit theorem for the number irreducible factors in the characteristic polynomial of a random  $T \in GL_n(F_q)$ . Professor Diaconis has remarked that this kind of result can be used as the basis for tests of random number generators.

For  $T \in GL_n(F_q)$ , let  $\Omega_n(T)$  be the number of irreducible factors (with multiplicity) that the characteristic polynomial of  $T$  has. Let  $\omega_n(T)$  be the number of *different* irreducible factors (without multiplicity). Let  $\mu_n := (1/\#GL_n(F_q)) \sum_T \omega_n(T)$  be the average number of different irreducible factors. Finally, let  $\sigma_n^2 := (1/\#GL_n(F_q)) \sum_T (\omega_n(T) - \mu_n)^2$  be the variance. Strong proved that  $\mu_n = \log n + O_q(1)$  and that  $\sigma_n^2 = \log n + O_q(1)$ . We prove the corresponding central limit theorem.

\* Research supported by the National Science Foundation (DMS-8901610).

† Supported by a Drexel University Faculty Development Minigrant.

**Theorem.** For any fixed  $x$ ,

$$\lim_{n \rightarrow \infty} P_n \left( \frac{\omega_n - \log n}{\sqrt{\log n}} < x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

*Remark.* It is slightly more convenient to work with  $\omega_n$  than with  $\Omega_n$ . Their difference has bounded mean and variance, so it is clear by a first or second moment argument that we get the same theorem for  $\Omega_n$ .

The proof requires Kung's "vector space cycle index." To state the result will require a bit of notation. If  $p$  is an irreducible polynomial over the finite field  $F_q$ , and if  $\lambda$  is an integer partition, define  $I_{p,\lambda}(T)$  to be 1 if  $p$  appears, associated with the partition  $\lambda$ , in the rational canonical form for  $T$ . Define it to be 0 otherwise. (The relevant linear algebra is concisely summarized in both [10] and [12]). The vector space cycle index is a polynomial in variables  $x_{p,\lambda}$  defined by

$$Z_n(q; \vec{x}) := \frac{1}{\#GL_n(F_q)} \sum_{T \in GL_n(F_q)} \prod_{p,\lambda} x_{p,\lambda}^{I_{p,\lambda}(T)}.$$

Let

$$G_q(u) := 1 + \sum_{n=1}^{\infty} Z_n(q; \vec{x}) u^n.$$

Finally, if  $\lambda = \{1^{b_1}, 2^{b_2}, \dots\}$  is the partition of  $j$  with  $b_i$  parts of size  $i$ , then let

$$d_j = b_1 + 2b_2 + \dots + (i-1)b_{i-1} + ib_i + \dots + ib_j.$$

The result we need is

**Lemma 1** (Kung/Stong).

$$G_q(u) = \prod_{p(z) \neq z} \left[ 1 + \sum_{j=1}^{\infty} \sum_{\lambda \vdash j} \frac{x_{p,\lambda}}{c_{\deg p}(\lambda)} u^{j \cdot \deg(p)} \right]$$

where  $p$  runs over monic irreducible polynomials,  $\deg p$  is the degree of  $p$ , and

$$c_m(\lambda) := \prod_i \prod_{k=1}^{b_i} (q^{md_i} - q^{m(d_i-k)})$$

To apply this theorem, one specializes; one makes appropriate substitutions for the variables  $x_{p,\lambda}$ . For example, in our case the probability generating function is obtained by setting  $x_{p,\lambda} := x$  for all  $p$  and  $\lambda$ :

$$\sum_{n,k} P_n(\omega_n = k) x^k u^n = \prod_{m \geq 1} (1 + x(\phi_m(u) - 1))^{e(m)}$$

where  $e(m)$  is the number of monic irreducibles of degree  $m$  for  $m > 1$  ( $e(1) := q - 1$ ), and

$$\phi_m(u) =$$

The last eq record for f

The "mom  $\mu_{n,r} := \sum_k k^r$ "

By a class fixed real

Let  $z_n = \prod_{m \geq 1} \phi_m$

For  $|u|$  analytic is,

By iso

$$\sum_{n=1}^{\infty}$$

Note

It r

$$\phi_m(u) = 1 + \sum_{j=1}^{\infty} \sum_{\lambda \geq j} \frac{u^{m\lambda}}{c_m(\lambda)} = 1 + \sum_{j=1}^{\infty} \frac{q^{m(j-1)} u^j}{(q^m - 1)(q^{m(j-1)} - 1) \cdots (q^m - 1)}$$

The last equality was proved in [12], together with the following one which we record for future use:

$$1 - \frac{1}{\phi_m(u)} = \sum_{j=1}^{\infty} \frac{(-1)^{j-1} u^{mj}}{(q^{mj} - 1)(q^{m(j-1)} - 1) \cdots (q^m - 1)}$$

The "moment generating function" is obtained by substituting  $e^t$  instead of  $x$ : if  $\mu_{n,r} := \sum_k k^r P_n(\omega_n = k)$ , and if  $M_n(t) = \sum_r (\mu_{n,r}/r!) t^r$ , then

$$\sum_n M_n(t) u^n = \prod_{m \geq 1} (1 + e^t (\phi_m(u) - 1))^{e(m)}$$

By a classical theorem in probability theory [4], it suffices to show that, for any fixed real number  $t$ ,

$$M_n(t/\sigma_n) \exp\left[\frac{-t\mu_n}{\sigma_n}\right] \rightarrow e^{t^2/2} \quad \text{as } n \rightarrow \infty.$$

Let  $z_n = (e^{t/\sigma_n} - 1)$ , and let  $g_n(u) = \prod_{m \geq 1} (1 - z_n (1 - \phi_m(u)))^{e(m)}$ . Note that  $\prod_{m \geq 1} \phi_m(u)^{e(m)} = 1/(1-u)$  (set  $t=0$  in the moment generating function). Hence

$$M_n\left(\frac{t}{\sigma_n}\right) = \text{COEFF}_{t^{-1}} \left\{ \frac{g_n(u)(1-u)^{-1}}{(1-u)^{z_n^{-1}}} \right\}.$$

For  $|u| < 1$ , let  $\delta_n(u) = \log g_n(u) - z_n \cdot \sum_{m \geq 1} u^m/m$ . Although  $\log g_n(u)$  is not analytic at  $u=1$ , we shall see that  $\delta_n(u)$  is the restriction of a function  $\Delta_n(u)$ , that is,

$$\delta_n(u) = \sum_{m=1}^{\infty} e(m) \sum_{l=1}^{\infty} \frac{z_n^l (-1)^{l-1}}{l} \left(1 - \frac{1}{\phi_m(u)}\right)^l - \sum_{m \geq 1} z_n \frac{u^m}{m}$$

By isolating the terms with  $l=1$ , we see that this is equal to

$$\sum_{m \geq 1} \left( e(m) z_n \left(1 - \frac{1}{\phi_m(u)}\right) - z_n \frac{u^m}{m} \right) + \sum_{m \geq 1} e(m) \sum_{l \geq 2} \frac{z_n^l (-1)^{l-1}}{l} \left(1 - \frac{1}{\phi_m(u)}\right)^l \\ = T_1 + T_2 \text{ say.}$$

Note first that  $e(m) = (q^m - 1)/m + (e(m) - (q^m - 1)/m)$ . Hence

$$T_1 = z_n \sum_{m \geq 1} \left( \frac{q^m - 1}{m} \left(1 - \frac{1}{\phi_m(u)}\right) - \frac{u^m}{m} \right) \\ + z_n \sum_{m \geq 1} \left( \left( e(m) - \frac{q^m - 1}{m} \right) \left(1 - \frac{1}{\phi_m(u)}\right) \right).$$

It is well known [11] that  $e(m) = (q^m - 1)/m + O(q^{m/2})$ . Furthermore,

$$1 - \frac{1}{\phi_m(u)} = \frac{u^m}{q^m - 1} + \sum_{j \geq 2} \frac{(-1)^{j-1} u^{mj}}{(q^{mj} - 1) \cdots (q^m - 1)}$$

$|E_n|$

Hence

By the Parse

$$T_1 = z_n \sum_{m \geq 1} \frac{q^m - 1}{m} \sum_{j \geq 2} \frac{(-1)^{j-1} u^{mj}}{(q^{mj} - 1) \cdots (q^m - 1)} + z_n \sum_{m \geq 1} O(q^{m-2}) \sum_{j \geq 1} \frac{(-1)^{j-1} u^{mj}}{(q^{mj} - 1) \cdots (q^m - 1)}$$

We have case where  $E$  show that  $E$  when  $t < 0$ . have

Now choose  $\rho$  so that  $1 < \rho < \sqrt{q}$ , say  $\rho := (1 + \sqrt{q})/2$ . Since the sums in  $T_1$  and  $T_2$  converge for  $|u| < \sqrt{q}$ , it is clear that  $\delta_n(u) = T_1 + T_2$  can be analytically continued to a function  $\Delta_n(u)$  that is analytic on the closed disc  $|u| \leq \rho$ . For future reference, observe that  $z_n \rightarrow 0$  as  $n \rightarrow \infty$ . The  $\Delta_n$ 's are uniformly bounded on  $|u| \leq \rho$ , and for all  $u$  on this disc we have  $\Delta_n(u) \rightarrow 0$  as  $n \rightarrow \infty$ .

On approach is motivated, in part, by a paper of Flajolet and Odlyzko [5] (see also [6] and [7]). Continuing where we left off, we have

We shall

$$\begin{aligned} M_n(t/\sigma_n) &= \text{COEFF}_{u^n} \left\{ \frac{g_n(u)(1-u)^{z_n}}{(1-u)^{z_n+1}} \right\} = \text{COEFF}_{u^n} \left\{ \frac{e^{\Delta_n(u)}}{(1-u)^{z_n+1}} \right\} \\ &= \text{COEFF}_{u^n} \left\{ \frac{e^{\Delta_n(1)}}{(1-u)^{z_n+1}} + \frac{(e^{\Delta_n(u)} - e^{\Delta_n(1)})}{(1-u)^{z_n+1}} \right\} \\ &= B_n + E_n \quad \text{say.} \end{aligned}$$

Observe

We must show that  $e^{-t\mu_n/\sigma_n} B_n + e^{-t\mu_n/\sigma_n} E_n = e^{t^2/2 + o(1)}$ . It is easy to estimate the major term:

Hence.

$$\begin{aligned} e^{-t\mu_n/\sigma_n} B_n &= e^{-t\mu_n/\sigma_n + \Delta_n(1)} \text{COEFF}_{u^n} \left\{ \frac{1}{(1-u)^{z_n+1}} \right\} \\ &= \frac{e^{-t\sqrt{\log n} + \Delta_n(1) - o(1)} \Gamma(n-1-z_n)}{\Gamma(1+z_n) \Gamma(n-1)}. \end{aligned}$$

If  $n$  is  $|u| < 1$  integr

Then Stirling's formula implies that

$$e^{-t\mu_n/\sigma_n} B_n = e^{-t\sqrt{\log n} - z_n \log n - \Delta_n(1) - o(1)} = e^{t^2/2 + o(1)}$$

For the error term, we consider separately the cases  $t > 0$  and  $t < 0$ . First suppose  $t > 0$ . To estimate  $E_n$ , take a circle of radius  $r_n = e^{-1/n}$  and apply Cauchy's theorem:

$$E_n := \frac{1}{2\pi i} \int_{|u|=r_n} \left( \frac{e^{\Delta_n(u)} - e^{\Delta_n(1)}}{1-u} \right) \frac{du}{(1-u)^{z_n+1}}$$

Inte:

Note that  $(e^{\Delta_n(u)} - e^{\Delta_n(1)})/(1-u)$  is uniformly bounded on the closed disc  $|u| \leq 1$ . Thus, for some  $n$ -independent  $K > 0$ , we have

$$|E_n| \leq \frac{K}{r_n^{n+1}} \int_{-\pi}^{\pi} \frac{d\theta}{|1 - r_n e^{i\theta}|^{z_n}} = \frac{K}{r_n^{n+1}} \int_{-\pi}^{\pi} |(1 - r_n e^{-i\theta})^{-z_n}| d\theta$$

By the Parseval-Bessel equality, this is equal to

$$\frac{2\pi K}{r_n^{n+1}} \sum_{k \geq 0} \left( r_n^k \frac{\Gamma(k + z_n/2)}{\Gamma(z_n/2)\Gamma(k + 1)} \right)^2 = O(1).$$

We have shown that, when  $t > 0$ , we have  $e^{-t\mu_n/\sigma_n} E_n = o(1)$ . Now consider the case where  $t < 0$ . The estimation is more delicate this time, because now we must show that  $E_n = o(e^{\sqrt{t \log n}})$ , not just  $O(1)$ . One reason that this can be done is that, when  $t < 0$ , we have  $z_n < 0$  (for  $n$  sufficiently large). For  $t < 0$  and  $0 < r < 1$ , we have

$$E_n := \frac{1}{2\pi i} \int_{|u|=r} \left( \frac{e^{\Delta_n(u)} - e^{\Delta_n(t)}}{1 - u} \right) \frac{du}{(1 - u)^{z_n} u^{n+1}}.$$

We shall re-express the integrand using Cauchy's theorem. For  $|u| \leq 1$  we have

$$e^{\Delta_n(u)} = \frac{1}{2\pi i} \int_{|w|=\rho} \frac{e^{\Delta_n(w)} dw}{(w - u)}.$$

Observe that  $1/(w - u) - 1/(w - 1) = (u - 1)/(w - u)(w - 1)$ . We therefore have

$$\frac{e^{\Delta_n(u)} - e^{\Delta_n(t)}}{1 - u} = \frac{-1}{2\pi i} \int_{|w|=\rho} \frac{e^{\Delta_n(w)} dw}{(w - u)(w - 1)}.$$

Hence, for  $0 < r < 1$ , we have

$$\begin{aligned} E &= \frac{1}{2\pi i} \int_{|u|=r} \left( \frac{-1}{2\pi i} \int_{|w|=\rho} \frac{e^{\Delta_n(w)} dw}{(w - u)(w - 1)} \right) \frac{(1 - u)^{-z_n} du}{u^{n+1}} \\ &= \frac{1}{(2\pi)^2} \int_{|w|=\rho} \frac{e^{\Delta_n(w)}}{(w - 1)} \int_{|u|=r} \frac{(1 - u)^{-z_n}}{(w - u)u^{n+1}} dudw. \end{aligned} \tag{1}$$

If  $n$  is large enough so that  $z_n < 0$ , then  $(1 - u)^{-z_n} = e^{-z_n \log(1 - u)}$  is analytic for  $|u| < 1$  and continuous for  $|u| = 1$ . Hence for the inner integral we can expand the integration path to a circle of radius 1:

$$\begin{aligned} \int_{|u|=r} \frac{(1 - u)^{-z_n}}{(w - u)u^{n+1}} du &= \int_{|u|=1} \frac{(1 - u)^{-z_n}}{(w - u)u^{n+1}} du \\ &= i \int_0^{2\pi} \frac{e^{-ni\theta} (1 - e^{i\theta})^{-z_n} d\theta}{(w - e^{i\theta})}. \end{aligned}$$

Integrating by parts, we get

$$-i \int_0^{2\pi} \frac{e^{-ni\theta}}{-ni} \left( \frac{d}{d\theta} \frac{(1 - e^{i\theta})^{-z_n}}{(w - e^{i\theta})} \right) d\theta = I_1 + I_2$$

ms in  $T_i$  and  
analytically  
 $|u| \leq \rho$ . For  
nly bounded

zko [5] (see

...}

estimate the

suppose  
Cauchy's

$|u| \leq 1$ .

where

$$I_1 := \frac{-z_n}{ni} \int_0^{2\pi} \frac{(1 - e^{i\theta})^{-z_n-1}}{(w - e^{i\theta})} e^{-(n-1)i\theta} d\theta$$

$$I_2 = \frac{i}{n} \int_0^{2\pi} \frac{(1 - e^{i\theta})^{-z_n}}{(w - e^{i\theta})^2} e^{-(n-1)i\theta} d\theta.$$

To estimate these integrals, we need three elementary facts. The first is that  $|1 - e^{i\theta}| = |2 \sin \theta/2|$ . The second is that, for  $\theta \in [0, \pi/2]$ ,  $\sin \theta \geq 2\theta/\pi$ . Finally, observe that, for  $|w| = \rho$ , we have  $|w - e^{i\theta}| \geq \rho - 1$ . With these facts we can bound  $I_1$  and  $I_2$ :

$$|I_1| \leq \frac{|z_n|}{n} \int_0^{2\pi} \frac{|(1 - e^{i\theta})^{-z_n-1}|}{|w - e^{i\theta}|} d\theta \leq \frac{|z_n|}{n(\rho - 1)} \int_0^{2\pi} \left| 2 \sin \frac{\theta}{2} \right|^{-z_n-1} d\theta$$

$$= \frac{|z_n| 2^{-z_n-1}}{n(\rho - 1)} \int_0^{\pi/2} \sin^{-z_n-1}(\theta) d\theta \leq \frac{|z_n| 2^{-z_n+1}}{n(\rho - 1)} \int_0^{\pi/2} \left( \frac{2\theta}{\pi} \right)^{-z_n-1} d\theta = O\left(\frac{1}{n}\right).$$

(For the last estimate, recall that  $z_n < 0$ .) By similar arguments,  $I_2 = O(1/n)$ . Thus we have bounded the inner integral of Eq. (1):

$$\int_{|u|=r} \frac{(1-u)^{-z_n} du}{(w-u)u^{n+1}} = O\left(\frac{1}{n}\right).$$

Now let  $M = \sup_n \max_{|w| \leq \rho} |e^{\Delta_n(w)}|$ . Then  $M < \infty$ , and we have

$$|E_n| \leq \frac{1}{4\pi^2} \int_{|w|=\rho} \frac{|e^{\Delta_n(w)}|}{|w-1|} \cdot \left| \int_{|u|=r} \frac{(1-u)^{-z_n}}{(w-u)u^{n+1}} du \right| |dw|$$

$$\leq \frac{1}{4\pi^2(\rho-1)} M \cdot O\left(\frac{1}{n}\right) 2\pi\rho$$

Thus we have  $|E_n| = O(1/n)$ , and consequently  $e^{-iu_n/\alpha_n} E_n \rightarrow 0$  as  $n \rightarrow \infty$ . ■

## REFERENCES

- [1] E. Bender, Central and local limit theorems applied to asymptotic enumeration, *J. Combin. Theory Ser. A*, **15**, 91-111 (1973).
- [2] E. R. Canfield, Central and local limit theorems for the coefficients of polynomials of binomial type, *J. Combin. Theory Ser. A*, **23**, 275-290 (1977).
- [3] E. Copson, *Asymptotic Expansions*, Cambridge University Press, Cambridge, 1965.
- [4] J. Curtiss, A note on the theory of moment generating functions, *Ann. Math. Stat.* **13**, 430-433 (1942).
- [5] P. Flajolet and A. Odlyzko, Singularity analysis of generating functions, *SIAM J. Discrete Math.*, to appear.
- [6] P. Flajolet and M. Soria, Gaussian limiting distributions for the number of components in combinatorial structures, *J. Combin. Theory Ser. A*, to appear.

- [7] P. Flajolet and M. Soria, General combinatorial schemes with gaussian and exponential tails and exponential tails, preprint.
- [8] M. Gerstenhaber, On the number of nilpotent matrices with coefficients in a finite field, *Ill. J. Math* 5(2), 330–333 (1961).
- [9] V. Goncharov, *AMS Transl.*, 19(2), 1–46 (1962).
- [10] J. Kung, The cycle structure of a linear transformation over a finite field. *Linear Algebra Appl.*, 36, 141–155 (1981).
- [11] F. MacWilliams and N. Sloane, *The Theory of Error Correcting Codes*. North-Holland, Amsterdam, 1981, p. 115.
- [12] R. Stong, Some asymptotic results on finite vector spaces. *Adv. Appl. Math.*, 9, 167–199 (1988).

Received April 18, 1990

It is that  
Finally,  
we can

$$\left(\frac{1}{n}\right).$$

$$O(1/n).$$

tion. *J.*

nials of

. 1965.

*Stat.* 13,

*AM J.*

σπρο-